Functional analysis of dsRNAs (L1, L3, L5, and M2) associated with isometric 34-nm virions of Agaricus bisporus (white button mushroom)

van der Lende, T.R.; Duitman, EH; Gunnewijk, MGW; Yu, L; Wessels, JGH

Published in:
Virology

DOI:
10.1006/viro.1996.0096

Publication date:
1996

VIROLOGY 217, 88–96 (1996)
ARTICLE NO. 0096

Functional Analysis of dsRNAs (L1, L3, L5, and M2) Associated with Isometric 34-nm Virions of *Agaricus bisporus* (White Button Mushroom)

TED R. VAN DER LENDE,[1] ERWIN H. DUITMAN, MARGA G. W. GUNNEWIJK, LI YU, and JOSEPH G. H. WESSELS

Department of Plant Biology, Groningen Biomolecular and Biotechnology Institute (GBB), University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

Received November 27, 1995; accepted December 26, 1995

cDNA clones of dsRNAs associated with La France disease of *Agaricus bisporus* were isolated. Clones corresponding to L1 and L5 dsRNAs were sequenced. The deduced amino acid sequence of L1 dsRNA (1078 amino acids, M_r 121K) showed significant homology with RNA-dependent RNA polymerases of other dsRNA viruses. The deduced amino acid sequence of L5 dsRNA (724 amino acids, M_r 82K) showed no homology with known proteins. Amino acid sequences of tryptic digests of three virion-associated proteins were determined. The 34-nm virion-associated protein of M_r 115K was encoded by the L1 dsRNA, thus identifying this protein as the RNA-dependent RNA polymerase. The virion-associated protein of M_r 90K was encoded by the previously sequenced L3 dsRNA. A cDNA clone of the previously sequenced M2 dsRNA was expressed in *Escherichia coli* and antibodies raised against this protein reacted only with a protein present in the cytoplasm of diseased *A. bisporus* fruit bodies but not in the 34-nm virions.

INTRODUCTION

Up to 10 major dsRNAs have been observed in fruit bodies and mycelium of the cultivated mushroom *Agaricus bisporus* affected by La France disease (Marino *et al.,* 1976; Hicks and Haughton, 1986; Ross *et al.,* 1986; Deahl *et al.,* 1987; Wach *et al.,* 1987; Harmsen *et al.,* 1989; Koons *et al.,* 1989; Morten and Hicks, 1992). In the Netherlands, 9 dsRNAs are associated with the disease: L1 (3.6 kb), L2 (3.0 kb), L3 (2.8 kb), L4 (2.7 kb), L5 (2.5 kb), M1 (1.6 kb), M2 (1.35 kb), S1 (0.86 kb), and S2 (0.78 kb), whereas L6 dsRNA can be found in both diseased and healthy fruit bodies (Harmsen *et al.,* 1989). The nucleotide sequences of L3, M1, and M2 dsRNAs have previously been reported (Harmsen *et al.,* 1991).

Fruit bodies from diseased crops also contain isometric particles of various diameters, 25- and 34-nm particles being the predominant types (Van Zaayen, 1979; Frost and Passmore, 1980). In addition, a bacilliform particle of 19 × 50 nm containing ssRNA was found (Tavantzis *et al.,* 1980, 1983; Revill *et al.,* 1994). At this point the only correlation is between the disease and the presence of the aforementioned nine, but sometimes six, dsRNA molecules (Harmsen *et al.,* 1989; Morten and Hicks, 1992).

Using a purification method that eliminates proteolysis to a large extent we have shown that the 34-nm particles purify together with the nine disease-associated dsRNAs and with three immunologically unrelated proteins of M_r 90K, 115K, and 120K (Van der Lende *et al.,* 1994). We prefer to call this particle ABV1 (*A. bisporus* virus 1) rather than the earlier designation LIV (La France isometric virus; Goodin *et al.,* 1992) because its causal relationship with the disease is still unproven.

In order to determine a relationship between the dsRNAs found in ABV1 and the constituent proteins of M_r 90K, 115K, and 120K (Van der Lende *et al.,* 1994), a cDNA library made of total dsRNA of *A. bisporus* was screened for cDNA clones corresponding to the disease-associated dsRNAs. In this paper we describe the nucleotide sequences of L1 and L5 dsRNA and identify L1 dsRNA as the genomic segment that encodes the RNA-dependent RNA polymerase of the ABV1 virus. Amino acid sequencing showed this to be the 115K virion-associated protein. The virion-associated protein of M_r 90K is encoded by the L3 dsRNA. The M2 dsRNA (Harmsen *et al.,* 1991) encodes a protein present in the cytoplasm of diseased *A. bisporus* fruit bodies but not in 34-nm virions.

MATERIALS AND METHODS

cDNA library

A cDNA library of dsRNAs from diseased mushrooms was constructed by Harmsen *et al.* (1991). To make this library, the dsRNAs were polyadenylated and priming with oligo$(dT)_{15}$ was used to synthesize cDNA. After second strand synthesis the double-stranded cDNA was inserted into the *Pst*I restriction site of pUC9 by G–C tailing.

Southern analysis

Southern blots were prepared from cDNA clones after digestion with *Pst*I. Blots were probed with randomly

[1] To whom correspondence and reprint requests should be addressed. Fax: +31 50 3632273; E-mail: T.R.van.der.Lende@Biol.rug.nl.

0042-6822/96 $18.00

primed cDNA fragments labeled with [α-^{32}P]dCTP prepared from individually isolated dsRNA molecules separated using 0.8% agarose gel electrophoresis. dsRNAs were isolated from diseased fruit bodies using the method of Morris and Dodds (1979).

RT–PCR

RT–PCR products were generated using primers based on partial sequenced cDNAs and with total dsRNAs as template. The reverse transcription reaction was performed using AMV reverse transcriptase (Life Technologies, Gibco BRL) following the manufacturer's protocol at 50° for 30 min. The PCR reaction was performed using Vent polymerase (New England's Biolabs) according to the manufacturer's direction but with the following adaptations: 1 μl of a 50-times diluted RT reaction mixture was used in a 50-μl PCR reaction mixture with the standard Vent polymerase buffer, using 30 cycles at 1 min at 90°, 2 min at 45°, and 1 min at 72°.

DNA sequencing

Specific fragments of L1 and L5 cDNAs were subcloned in pUC vectors for DNA sequencing. Additional cDNAs were constructed by RT–PCR using primers based on already determined sequences and total dsRNAs prepared from diseased *A. bisporus* fruit bodies. Both strands of the cDNAs were sequenced using a T7 DNA polymerase kit (Pharmacia) according to the dideoxy chain termination method (Sanger *et al.,* 1977). Deduced amino acid sequences were compared with the EMBL and SWPROT databases using the FASTP program (Lipman and Pearson, 1985). Multiple alignments were performed with the CLUSTAL program (Higgins and Sharp, 1988). The L1 dsRNA sequence is available from the EMBL database as Accession No. X94361; for L5 dsRNA the number is X94362.

Amino acid sequencing

Virion isolation and electrophoresis of proteins were as described by Van der Lende *et al.* (1994). Protein fragments of the virion-associated proteins were isolated by high pressure liquid chromatography (HPLC) of tryptic digests of individual virion-associated protein bands cut from a standard 7.5% SDS–PAGE gel. Purified fragments were sequenced using Edman degradation and a pulse liquid sequenator on line connected to a PTH analyzer (Applied Biosystems, U.S.A.).

Antibodies against M2 protein produced in *Escherichia coli*

A 512-bp *Bcl*I/*Pst*I restriction fragment from the full-length cDNA clone of M2 dsRNA (pAV322; Harmsen *et al.,* 1991) was expressed in *E. coli* using the plasmid pQE9 and the *E. coli* strain M15, containing the repressor plasmid pREP4 (QIAexpressionist; QIAGEN, U.S.A.) according to the manufacturer's protocol. The expressed heterologous protein was directly cut from the gel after SDS–PAGE and electroeluted for antibody production. Rabbits were subcutaneously injected five times at 10-day intervals with the electroeluted protein in Freund's complete adjuvant (1/1 v/v). Nonspecific antibodies were removed by incubating the antisera with proteins extracted from healthy *A. bisporus* fruit bodies and which were immobilized on polyvinylidene difluoride membranes (Immobilon-P; Millipore Corp., U.S.A.).

RESULTS

Isolation of virus-specific cDNA clones

A number of cDNA clones were isolated from the cDNA library constructed by Harmsen *et al.* (1991) and identified by hybridization with individually isolated and [α-^{32}P]dCTP-labeled dsRNAs. Thirty-seven cDNAs were isolated corresponding to disease-related dsRNAs. None of these cDNAs appeared to be full-length but together the clones could represent all sequences present because 3′ ends of both strands of the dsRNAs were used in the initial polyadenylation reaction.

For L5 dsRNA two specific cDNA clones called 10.3 (approx 1.5 kb) and 12.36 (approx 1 kb) were isolated, which did not hybridize to each other and were probably generated from opposite ends. The estimated lengths of L5 dsRNA and the combined lengths of the two cDNA clones (2.5 kb) suggested that virtually the complete dsRNA was cloned. Because the clones did not overlap, an additional cDNA clone was generated using dsRNA as template in an RT–PCR reaction with primers based on determined sequences in these partial cDNA clones. In Fig. 1A the original cDNA clones and the RT–PCR-derived cDNA clone (L5pcr1), which were used for subcloning and subsequent nucleotide sequencing, are schematically presented.

Restriction enzyme mapping and Southern blot analysis of clones specific for L1 dsRNA revealed a more complex situation. Apparently the initial reverse transcription reaction using oligo(dT) primers (Harmsen *et al.,* 1991) had also occurred upstream of the added poly(A) tails, probably at A-rich stretches within the L1 nucleotide sequence. Some cDNAs did not contain AT-rich stretches and could represent cDNAs which were fragmented during synthesis and subsequently cloned into a *Pst*I site by G–C tailing. Two RT–PCR-derived cDNA clones (L1pcr1 and L1pcr2) were used to obtain preliminary sequence data and to support the mapping of the clones. In Fig. 1B the original five cDNA clones which were used for subcloning and subsequent nucleotide sequencing and the two RT–PCR-derived cDNA clones are schematically presented.

Using the clones shown in Figs. 1A and 1B the complete sequences of L5 dsRNA (2455 bp) could be determined. For the L1 dsRNA 3396 of the approximately 3600 bp were determined.

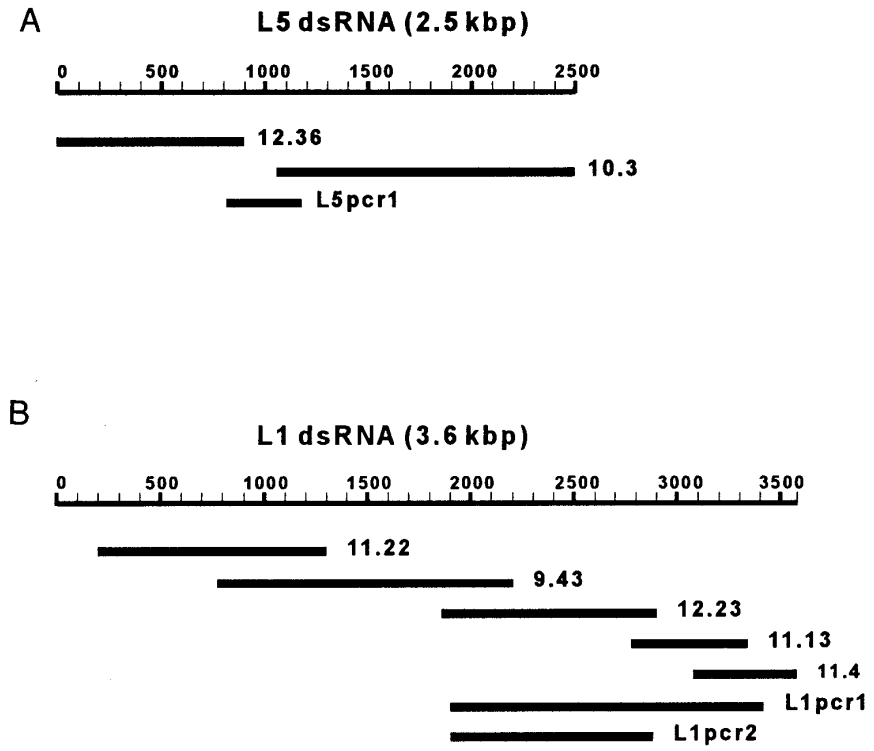

FIG. 1. Schematic presentation of original cDNA clones (indicated by numbers) for (A) L5 and (B) L1 dsRNA isolated from the cDNA library constructed by Harmsen *et al.* (1991). The positions of clones are drawn with respect to the complete dsRNAs, indicating internal initiation of cDNA synthesis for the L1 dsRNAs. Additional cDNAs derived using RT–PCR are also indicated (L1pcr1, L1pcr2, and L5pcr1). The first 200 bp of L1 dsRNA were not found in the cDNA library.

Nucleotide sequences of the L1 and L5 dsRNAs

The nucleotide sequences of L1 and L5 dsRNAs and the predicted amino acid sequences are shown in Figs. 2 and 3, respectively. The 3396-bp sequence derived from the L1 dsRNA has a base composition of the coding (+) strand of A (31%), C (15%), G (25%), and U (29%). L5 dsRNA consists of 2455 bp with a base composition of the coding (+) strand of A (34%), C (15%), G (26%), and U (25%). A stretch of 23 nt (GGCAACGGCUAGUUGGCCAAAUU), underlined in Fig. 3, has also been found in the 5′ region of the (+) strand of the M2 dsRNA (Harmsen *et al.,* 1991), whereas part of this sequence (AACGGCUAGUU, double underlined in Fig. 3) is also present in the 5′ region of the (+) strand of L3 dsRNA (Harmsen *et al.,* 1991).

Open reading frame analysis

L1 dsRNA contains one large open reading frame (ORF) of 1078 amino acids starting at nucleotide 13 with the codon AUG and ending at nucleotide 3249 with the codon TAA. This ORF encodes a putative protein of M_r 121,823 (Fig. 2). However, because 200 bp upstream of this ORF were not determined, this protein could be about 60 amino acids larger. A number of 20 smaller open reading frames ranging from 31 to 114 amino acids were detected, encoding putative proteins ranging from M_r 3724 to 13,492. The largest of these ORFs are encoded by the strand complementary to the strand encoding the major ORF.

L5 dsRNA contains one large open reading frame of 724 amino acids starting at nucleotide 125 with the codon AUG and ending at nucleotide 2299 with codon TAA. This frame encodes a putative protein of 724 amino acids of M_r 81,912 (Fig. 3). Five smaller open reading frames were detected ranging from 35 to 98 amino acids encoding putative proteins ranging from M_r 4106 to 11,442. Also in this case the largest ORFs are encoded by the strand complementary to the strand encoding the major ORF.

Codon usage

Analysis of the codons used by the two dsRNAs in the large open reading frames presented above and in the three dsRNAs sequenced by Harmsen *et al.* (1991) revealed that the codon usage of the viral genes is biased.

FIG. 2. Nucleotide sequence of 3396 nt of the (+) strand of L1 dsRNA (3600 bp) and the amino acid sequence of the longest ORF. Numbering of the nucleotides is with respect to the first determined nt (probably bp 200 of the dsRNA molecule). Numbering of the deduced amino acids is with respect to the initiation codon (AUG). The shaded sequence represents the region with homology to RNA-dependent RNA polymerases of dsRNA viruses (see Fig. 4).

```
   1 ACUUGUUUGAAUAUGUUGAGUGGAUAUGAUAGUGACCAGAUAAUUGGGUGGGCUGAGGAUAUUCCAAAUUUUGAUGUUGAGUAUGCUUGCAUUUCACACA
   1             M  L  S  G  Y  D  S  D  Q  I  I  G  W  A  E  D  I  P  N  F  D  V  E  Y  A  C  I  S  H

 101 CACGUGCGAAAGCAGUGGUUGGUGAGGACAUUGCUCAUAAUUUGUCCUGCUAUUCAAGGUUACGUACCGCAUUGUCUAAUAUAAACCCAGACGAUAAUGC
  30 T  R  A  K  A  V  V  G  E  D  I  A  H  N  L  S  C  Y  S  R  L  R  T  A  L  S  N  I  N  P  D  D  N  A

 201 GCCUUUCAUUUUUGUCCAAACUUGUGAACUAGCAGAAGCAAUUGGAGUACAGAUUGCGAUGGUAUUGAUAGUCAAGGAGGGGAUAAUUCAACAAACUGAU
  64   P  F  I  F  V  Q  T  C  E  L  A  E  A  I  G  V  Q  I  A  M  V  L  I  V  K  E  G  I  I  Q  Q  T  D

 301 AGGUAUAAAAGUUUGUCAAGAGAAGAACAGAAAACUCUUAAACAUGUGCAUGAUGUUCAAAAUAAUCAUAAUAAAUUGUGGGUAGAAAAAUAUAAGUAUA
  97 R  Y  K  S  L  S  R  E  E  Q  K  T  L  K  H  V  H  D  V  Q  N  N  H  N  K  L  W  V  E  K  Y  K  Y

 401 CGCAUAGGGUUGUCUUAUCUUAUAAUGAUAUUUACCACAAGGUUAAGGGAUUAUAUAUUAGUGUUGGGUGUGUCAAAAGCUCUUUUGGUACAGACGAUGU
 130 T  H  R  V  V  L  S  Y  N  D  I  Y  H  K  V  K  G  L  Y  I  S  V  G  C  V  K  S  S  F  G  T  D  D  V

 501 GCUGAGUGGGAUAGAUGAGUACACUUAUGAUGCUCGCGAGAAGGUAAAUGAGGUUUCGAAUUUGCUUAAAGAGUCGAAGCAUCCGGCGACCAAAGCUGUA
 164   L  S  G  I  D  E  Y  T  Y  D  A  R  E  K  V  N  E  V  S  N  L  L  K  E  S  K  H  P  A  T  K  A  V

 601 UGCGCGAGGUAUUUGCAGCAAAAUUUUCAGGCGGCAGCGCCAGACUAUGCUAAUGCAAGUUGCUCCCACGUUGCGUGGUCGAGAUGGGUGAAUGCAGUCA
 197 C  A  R  Y  L  Q  Q  N  F  Q  A  A  A  P  D  Y  A  N  A  S  C  S  H  V  A  W  S  R  W  V  N  A  V

 701 AAUCGCACGUCAGGACUGAAAGCGCGCAUUAUCCAACAAAUAAUAUGGUGAUGGCCAUGUGUGACGAGGAGUAUGUGAUAGAGAAUUUUCCUCAUCAAGC
 230 K  S  H  V  R  T  E  S  A  H  Y  P  T  N  N  M  V  M  A  M  C  D  E  E  Y  V  I  E  N  F  P  H  Q  A

 801 AGGGUCAGAUAAGUUUGCUGCCUGUCAUAUAAGUACAGUGAUCAAGGAUCACAUGGAUAAGUUGAAAAAACCAACGAAUGGUAGACCUGUGUUUGUUCAG
 264   G  S  D  K  F  A  A  C  H  I  S  T  V  I  K  D  H  M  D  K  L  K  K  P  T  N  G  R  P  V  F  V  Q

 901 ACUGCAUGUUUGAUUGGUUCGGUCUUGAAGAAUAAAGAUAUCCAGUGCUCGUAUGAGAAGCUAUUGUCUUUUAUAUUAUAUGAAAUGUUCUUGUACACAA
 297 T  A  C  L  I  G  S  V  L  K  N  K  D  I  Q  C  S  Y  E  K  L  L  S  F  I  L  Y  E  M  F  L  Y  T

1001 CGGAUGUUGAGAUGCAUAAAUUGUGGAGUAGUGUGGUAUAUUCGCCUGUAGAUAAAGAUACUUUCUUUUUGUGUGCAGAGUAUGCUCAUGCGAGGAUAAG
 330 T  D  V  E  M  H  K  L  W  S  S  V  V  Y  S  P  V  D  K  D  T  F  F  L  C  A  E  Y  A  H  A  R  I  R

1101 GACAUCAGGAACGUAUUUAGGCAUAAAAUUGACUGGCCGUCCAACAGAUGUCAUAUAUGUUGUAUCCGCAAUGUUUAGCCUGGUAGAGUGUAUGGCAAAG
 364   T  S  G  T  Y  L  G  I  K  L  T  G  R  P  T  D  V  I  Y  V  V  S  A  M  F  S  L  V  E  C  M  A  K

1201 AGGAUUUGGCACAAGAGUUUGACGAUCGUACUUCUGAUAGGCCGGCGAAGAGGUUCUGGUCUAAUGGAGUGCUUAAGCGAAGAAGAAUAUAAUAUAAGGU
 397 R  I  W  H  K  S  L  T  I  V  L  L  I  G  R  R  R  G  S  G  L  M  E  C  L  S  E  E  E  Y  N  I  R

1301 UCAAUAAUGCGAUACUGAACUGUUAUUCGCAUAUGACAACUGGAUAUAGUGCUAAGUUUGGGCGGUUGGCGAAUGAUGUUGAAUCGUUCAAAUCGUUUAU
 430 F  N  N  A  I  L  N  C  Y  S  H  M  T  T  G  Y  S  A  K  F  G  R  L  A  N  D  V  E  S  F  K  S  F  M

1401 GCAGUUGCGAAAGCCUGGUCACAGCCAGGUUCCCGGCAACCGGUGCACGAAAGUUAUUAUUAGAUUGACGGAACAUAUGGCUGAAUUAGAUGCUGAUCCU
 464   Q  L  R  K  P  G  H  S  Q  V  P  G  N  R  C  T  K  V  I  I  R  L  T  E  H  M  A  E  L  D  A  D  P

1501 UCAGCUGGUGAGUUAUUGACUGAUCCCAUCGAUAGGAUUCAAUUGAGACUUAACAAGUCAACAUUGUUGAAUUUUCUGCCCGUAGUAGAAGCGGUGGAAG
 497 S  A  G  E  L  L  T  D  P  I  D  R  I  Q  L  R  L  N  K  S  T  L  L  N  F  L  P  V  V  E  A  V  E

1601 CUGCCUGGCGGAAUUAUGAUCCGAAUAGUUUUACUGGUGUGUUCUGGAAACAUGAAGUAGGGAAAAACGCAUCUCGAUCAUUGUGGCCAGCGCAUCUAGU
 530 A  A  W  R  N  Y  D  P  N  S  F  T  G  V  F  W  K  H  E  V  G  K  N  A  S  R  S  L  W  P  A  H  L  V

1701 CCAUUACGUUCUAGUGUCCAUGAUACUGCACUUAAUCGACAAGUCAGGAGAAAUACCAGGAUCACGUAACAAUGCACCUAGUGAUAGGCAAUUGAAGGAU
 564   H  Y  V  L  V  S  M  I  L  H  L  I  D  K  S  G  E  I  P  G  S  R  N  N  A  P  S  D  R  Q  L  K  D

1801 CACUGGAUGUGGUCAGAAUGCCAAGAUUUUGUGCCUUUAAUGAUGGAUUACGCUAAUUUUAAUGAACAGCAUAGCAUAGAGGCGAUGAAAGCAACGAUAC
 597 H  W  M  W  S  E  C  Q  D  F  V  P  L  M  M  D  Y  A  N  F  N  E  Q  H  S  I  E  A  M  K  A  T  I

1901 AACCACUGCGCGAUGUGUAUGCUAAGGCAGGAGUGUUGUCUAAGGAUCUAGCUGAUGCUAUUGAUUGGGUAGUCAAGUCGUUUGAUCAAAUAUGCGCUAU
 630 Q  P  L  R  D  V  Y  A  K  A  G  V  L  S  K  D  L  A  D  A  I  D  W  V  V  K  S  F  D  Q  I  C  A  I

2001 AGACGAAAAUGGUGACUUGAGGCGAUUUACACACGGCCUGCUAUCUGGAUGGAGAUGUACGGCUUACAUCAAUAAUUUAAUAAAAUAUAGCUCAGUACGAA
 664   D  E  N  G  D  L  R  R  F  T  H  G  L  L  S  G  W  R  C  T  A  Y  I  N  N  L  I  N  I  A  Q  Y  E

2101 GUGGGUAGACAACAGCUGCAUGAGUUUUUUGGAAUGAGUUCUUCAUACCAGUUCGACACAGGAGGAGAUGAUGGAUGUGCAGAUGAGAAAGACUUGGUGA
 687 V  G  R  Q  Q  L  H  E  F  F  G  M  S  S  S  Y  Q  F  D  T  G  G  D  D  G  C  A  D  E  K  D  L  V

2201 CAGCUUAUUCGUUGUUGCGUAGCAUGACUGCUAUGGGAUUUGAGUUUAAGGAUAUAAAACAAUUAAUAAGCAGCGGAACGCACGAAUUUUUUAGAUUGUU
 730 T  A  Y  S  L  L  R  S  N  T  A  M  G  F  E  F  K  D  I  K  Q  L  I  S  S  G  T  H  E  F  F  R  L  L

2301 AAUAACUCCAGAAGGUGUAUUCGGUUCAGUGAUCAGAAUGUUAGCGUCUGCUGCAUCAGGCCAAUGGUCCAAUUCAGUGAGAGCGAAAUUAGUGGAUCCA
 764   I  T  P  E  G  V  F  G  S  V  I  R  M  L  A  S  A  A  S  G  Q  W  S  N  S  V  R  A  K  L  V  D  P

2401 CUUUCAAAGAUGACUAGUAUAAUGGACAUAAAACAUAAAUUAUGGAGGAGAUCGGGUUACAAUGAUGAUUGGGCUGAACAAUUGCAACAUUAUAUGAGCC
 797 L  S  K  M  T  S  I  M  D  I  K  H  K  L  W  R  R  S  G  Y  N  D  D  W  A  E  Q  L  Q  H  Y  M  S

2501 UGAAAUGGGUAGAUCCGGAGAGCCCAAAGGAUUUGACAAAUUUUGUGCAUGGCCAGAAAAGCACGGGAGGUUUUUUGGGCAUACCAGACUGUCAAGGUAG
 830 L  K  W  V  D  P  E  S  P  K  D  L  T  N  F  V  H  G  Q  K  S  T  G  G  F  L  G  I  P  D  C  Q  G  R

2601 GCUAUAUGAGUUAGCGAGUGUGAGUGGAAUAGAAAAGCCAUCAAAAAUAGAAUUGCGCUCAUUCCCGCACGAUGCGAGUGAGCAGUUGCUGAAUAAAUUG
 864   L  Y  E  L  A  S  V  S  G  I  E  K  P  S  K  I  E  L  R  S  F  P  H  D  A  S  E  Q  L  L  N  K  L

2701 GCUCCAGACAUAGAACAGAUGGUCGGUGCUGAUCAGAUGGAAGAUAUGGAACGUAUAGCUGUGAACAUGAGUAAAAUGGUAUUUGUUGGUAACGUAGCUU
 897 A  P  D  I  E  Q  M  V  G  A  D  Q  M  E  D  M  E  R  I  A  V  N  M  S  K  M  V  F  V  G  N  V  A

2801 CAGGAUUUAGUCCGGCUAUAGCAAGUAAAAUGGUAGAAUCUACAAAUUUGAGAAAAGUACGAGCAAAAAGGACAGGCAGGGUAUUUGCUGUCGAUAAUAC
 930 S  G  F  S  P  A  I  A  S  K  M  V  E  S  T  N  L  R  K  V  R  A  K  R  T  G  R  V  F  A  V  D  N  T

2901 AUCGCCUUUAACGAUAGAUGAUGUUAGGAAAGAUCUUGAUUUAUGGUACCCUACGAUUGAAGAUAUCAAGAAGGUAAUUGGUGACUAUAGUUCGAUGAGU
 964   S  P  L  T  I  D  D  V  R  K  D  L  D  L  W  Y  P  T  I  E  D  I  K  K  V  I  G  D  Y  S  S  M  S

3001 CUUUUGACUAAACCAACAAGUCAUGAAACUUUAUUGAUGAAGUUAAGCGUCAAACAUAAUGUAAAAUUAUACAGUAGGUUACAUUAUAUGUUGAUGGCUG
 997 L  L  T  K  P  T  S  H  E  T  L  L  M  K  L  S  V  K  H  N  V  K  L  Y  S  R  L  H  Y  M  L  M  A

3101 AACCAGACAUAACAGGUCUGGGACCAAUAUUGGCAACUGAAGAUUAUUACAAGGAUCUUUUGAUAUUGUCUUUUCUGACUGCAGAAAAACUGAAUGCUAG
1030 E  P  D  I  T  G  L  G  P  I  L  A  T  E  D  Y  Y  K  D  L  L  I  L  S  F  L  T  A  E  K  L  N  A  R

3201 AGGAGUAAGUCAAAAGUUCUCAGAUUAUGCAGUAGAUGGCAAUGGGUAAAAGGAUGAAUUACUAAAGCAAGCAAUAAAGAAUGGAACAACCAAAUGUACG
1064   G  V  S  Q  K  F  S  D  Y  A  V  D  G  N  G

3301 UAGCGUGCUGGUUGUUGGUGAAAUGAUCGAAUCAUACGUAGUGUUUCGAUCGUUAAAUUGUUUAAGUGAACACGUGUAGUGUGCAACUUAAAUGGC
```

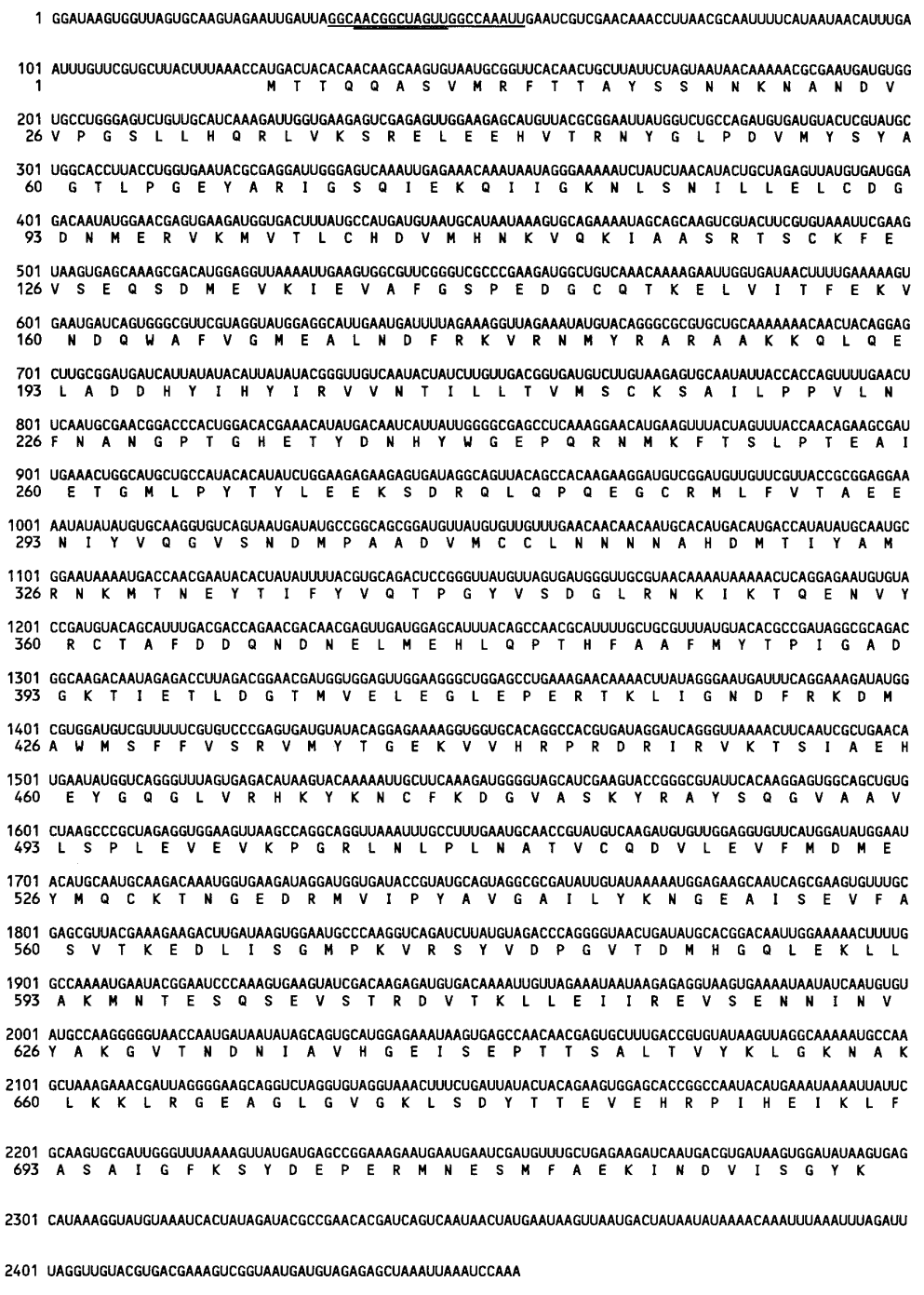

FIG. 3. Nucleotide sequence of the (+) strand of L5 dsRNA and the amino acid sequence of the longest ORF. Numbering of the nucleotides is with respect to the first determined nucleotide. Numbering of the deduced amino acids is with respect to the initiation codon (AUG). A stretch of 23 nt (GGCAACGGCUAGUUGGCCAAAUU) which is underlined is also present in the 5′ region of the (+) strand of M2 dsRNA; part of this sequence (AACGGCUAGUU, double underlined) is also present in the (+) strand of L3 dsRNA (Harmsen *et al.,* 1991).

Table 1(b) shows that, in contrast to the ABV1 genes, the *A. bisporus* genes have a strong bias toward C or T at the third position of each codon, particularly in highly expressed genes.

When individual codons are examined it is clear that the codon usage for almost every amino acid in the putative proteins of ABV1 is completely different from that of the host genes (Table 1a).

L1 dsRNA encodes a putative RNA-dependent RNA polymerase

The putative protein encoded by the major ORF of L1 dsRNA revealed significant homology with consensus sequences of RNA-dependent RNA polymerases (RdRp) of (+) strand RNA and dsRNA viruses, although the overall similarity is low. Several authors have recognized con-

TABLE 1

(a) Overall Codon Usage of the Major ORFs of L1, L3, L5, M1, and M2, Comprising 3294 Codons of *Agaricus bisporus* Virus

		ABV1	MBV	A.b			ABV1	MBV	A.b.			ABV1	MBV	A.b			ABV1	MBV	A.b
TTT	Phe	66	44	21	TCT	Ser	16	23	24	TAT	Tyr	70	59	34	TGT	Cys	64	68	33
TTC	Phe	34	56	79	TCC	Ser	4	12	27	TAC	Tyr	31	42	66	TGC	Cys	36	32	67
TTA	Leu	23	14	2	TCA	Ser	18	16	13	TAA	---	60	50	17	TGA	---	20	25	17
TTG	Leu	43	25	18	TCG	Ser	18	23	11	TAG	---	20	25	67	TGG	Trp	100	100	100
CTT	Leu	9	24	20	CCT	Pro	27	36	34	CAT	His	76	65	52	CGT	Arg	14	22	29
CTC	Leu	1	13	40	CCC	Pro	10	26	34	CAC	His	24	35	48	CGC	Arg	9	13	23
CTA	Leu	11	11	5	CCA	Pro	41	18	18	CAA	Gln	59	52	58	CGA	Arg	13	19	8
CTG	Leu	14	13	16	CCG	Pro	23	19	14	CAG	Gln	41	48	42	CGG	Arg	12	11	5
ATT	Ile	28	49	36	ACT	Thr	34	42	39	AAT	Asn	74	59	30	AGT	Ser	29	16	12
ATC	Ile	17	31	58	ACC	Thr	10	11	38	AAC	Asn	26	41	70	AGC	Ser	15	11	13
ATA	Ile	55	20	6	ACA	Thr	37	20	14	AAA	Lys	54	60	39	AGA	Arg	23	20	12
ATG	MET	100	100	100	ACG	Thr	19	27	10	AAG	Lys	46	41	61	AGG	Arg	30	14	24
GTT	Val	24	27	32	GCT	Ala	30	45	42	GAT	Asp	78	69	52	GGT	Gly	33	47	38
GTC	Val	11	24	47	GCC	Ala	6	14	34	GAC	Asp	22	31	48	GGC	Gly	15	14	31
GTA	Val	30	24	9	GCA	Ala	39	25	14	GAA	Glu	62	51	62	GGA	Gly	34	29	24
GTG	Val	36	25	12	GCG	Ala	26	17	10	GAG	Glu	38	49	38	GGG	Gly	18	10	8

Note. The codon usage of the four major ORFs (comprising 1458 codons) of mushroom bacilliform virus (MBV) (Revill *et al.,* 1994) and the codon usage of *Agaricus bisporus* genes based on a table by Van der Vlugt *et al.* (1993) (comprising 2552 codons) are included. Codon usage is given as percentage for each amino acid.

(b) Base Preference at the Third Position of Each Codon, for ABV1, MBV, *Agaricus bisporus* Genes and Its Highly Expressed GPD$^{ag\text{-}2}$ Gene (Harmsen *et al.,* 1992) and ABH1 Hydrophobin Gene (L. G. Lugones and J. G. H. Wessels, Unpublished)

Pos 3	ABV1	MBV	A.b	GPD	ABH1
T	30	33	30	32	37
C	12	19	38	40	47
A	30	23	15	9	7
G	28	25	17	19	9

served regions in the RdRp's which have been sequenced so far (Argos, 1988; Poch *et al.,* 1989; Koonin and Dolja, 1993). Ribas and Wickner (1992) defined essential domains of RdRp's of the L-A virus of *Saccharomyces cerevisiae* by mutagenesis analysis of the most conserved regions. In Fig. 4 the boxed amino acids represent the consensus sequence for RdRp's of dsRNA viruses as indicated by Ribas and Wickner (1992). Bruenn (1993) aligned a number of RdRp's of dsRNA viruses found in fungi and protozoans which showed homology beyond this consensus. These homologies were also found in the RdRp of ABV1 (Fig. 4).

Amino acid sequencing of ABV1 proteins

N-terminal sequencing of ABV1 proteins isolated as described by Van der Lende *et al.* (1994) was not possible because the proteins were entirely blocked at their N-

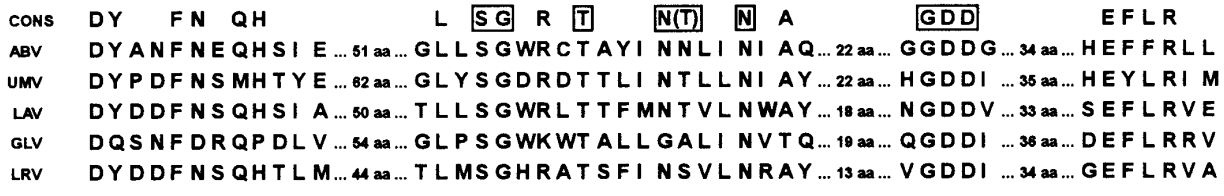

FIG. 4. Alignment of the amino acid sequence of ABV1 L1 dsRNA shaded in Fig. 3 with amino acids considered to be in the consensus sequences of RNA-dependent RNA polymerases of other dsRNA viruses. ABV, *A. bisporus* virus 1; LAV, L-A virus of *S. cerevisiae* (Icho and Wickner, 1989); UMV, *Ustilago maydis* virus found in the fungus *U. maydis* (Bruenn, 1993); GLV, *Giardia lamblia* virus found in the protozoan *G. lamblia* (Wang and Wang, 1986; White and Wang, 1990; Furfine and Wang, 1990); LRV, *Leishmania* RNA virus found in the protozoan *L. braziliensis* subsp. *guyanensis* (Widmer *et al.,* 1989; Widmer and Patterson, 1991). Amino acids which are conserved among these five viruses are indicated at the top as CONS of which the sequences which are considered to be the consensus sequences of RNA-dependent RNA polymerases of dsRNA viruses are boxed. The position of this region, at the center of the RdRp protein, is similar for all of these RdRp's (Bruenn, 1993).

termini. Individual viral protein bands, isolated from a standard 7.5% SDS–PAGE gel, were therefore subjected to tryptic digestion and protein fragments were isolated by HPLC.

Three different HPLC fractions of the digested 90K protein were selected for sequence analysis. The first sequence found was NAAGDAYVIRETITV. This sequence is nearly identical to the last 15 C-terminal amino acids deduced from the nucleotide sequence of L3 dsRNA (amino acids 774 to 788; Harmsen *et al.,* 1991) except for the underlined amino acid valine (V) which appeared an isoleucine (I) residue in the amino acid sequence deduced by Harmsen *et al.* (1991). The amino acid sequences of fragments present in the two other HPLC fractions were LPTLPV and TSXVVV (where X indicates an unidentified amino acid). These sequences can also be found in the amino acid sequence deduced from L3 dsRNA, specifically amino acids 592 to 597 and amino acids 746 to 751, respectively, while the X was an aspartate residue (D) in the sequence deduced by Harmsen *et al.* (1991).

A fraction derived from the digested 115K protein band yielded LYELASVSG/DIEKPSKIELR/S with ambiguities at two positions. This sequence was found at the amino acid positions 864 to 882 in the amino acid sequence deduced from L1 dsRNA as presented in Fig. 2, with a G and an R at the ambiguous positions.

A fraction isolated from the fragmented 120K proteins yielded the amino acid sequence VAGTGTIEFGDGLEAGASWPEAPER. This sequence did not correspond to amino acid sequences deduced from sequenced ABV1 dsRNAs nor to any sequence in protein databases. *In vitro* translation of denatured L2 dsRNAs showed a translation product of M_r 120K, suggesting that the largest virion-associated protein of M_r 120K is encoded by the L2 dsRNA but this remains conjectural in the absence of sequence data.

Identification of the protein encoded by M2 dsRNA

A fragment of 512 bp, cut from the full-length cDNA clone of M2 dsRNA (pAV322; Harmsen *et al.,* 1991) with *Bcl*I and *Pst*I was cloned in the vector pQE9. This fragment encodes 170 of the 340 amino acids (from amino acid 120 to 290) of the putative protein encoded by the major ORF of M2 dsRNA (Harmsen *et al.,* 1991). *E. coli* transformed with this construct was induced to express the viral cDNA insert. An extra protein band at about M_r 18K could be observed in *E. coli* extracts after standard SDS–PAGE and staining of the gel with Coomassie brilliant blue (Fig. 5A). The apparent molecular weight of this band was in agreement with the expected M_r of 19K. The protein of M_r 18K, putatively containing a partial amino acid sequence encoded in M2 dsRNA, was cut from the gel, electroeluted, and injected into rabbits. The resulting antiserum, purified against *E. coli* proteins, did not react with Western blots of virion proteins (Van der Lende *et al.,* 1994) but a protein of about M_r 40K could be detected on Western blots of total protein preparations from diseased *A. bisporus* fruit bodies (Fig. 5B, lane D). This molecular weight is in agreement with that deduced from the nucleotide sequence of M2 dsRNA (Harmsen *et al.,* 1991). On Western blots of total proteins from healthy *A. bisporus* fruit bodies this protein was not observed (Fig. 5A, lane H).

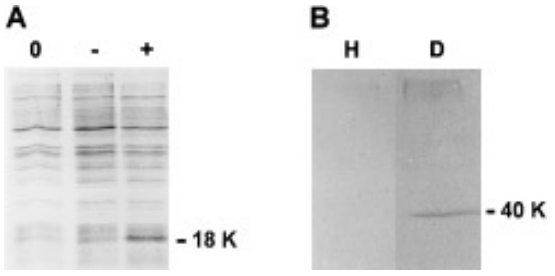

FIG. 5. (A) Expression of a 512-bp fragment, cut from the full-length cDNA clone of M2 dsRNA (pAV322) and inserted in expression vector pQE9, in *E. coli* M15. Shown are total protein extracts after standard 12.5% SDS–PAGE and staining with Coomassie brilliant blue. Lane 0, untransformed cells after induction with IPTG. Cells carrying the construct pQE9 with the viral insert are shown without (lane −) and after (lane +) induction. An extra protein of about M_r 18K could be observed after induction with IPTG. (B) Detection of a protein of about M_r 40K on Western blots of total protein extracts from diseased *A. bisporus* fruit bodies (lane D). Extracts from healthy *A. bisporus* fruit bodies showed no reaction (lane H).

DISCUSSION

The main objective of this research was to demonstrate the relationships between the virus-encoded proteins and the dsRNAs contained in the 34-nm virions of *A. bisporus.* Two major virion-associated proteins of M_r 120K and 90K in addition to a minor protein of M_r 115K were described (Van der Lende *et al.,* 1994). The 34-nm virions copurified with six different dsRNA molecules (L1, L2, L3, L4, L5, and M2) sometimes accompanied by three additional dsRNAs (M1, S1, and S2) (Van der Lende *et al.,* 1994; Goodin *et al.,* 1992). These dsRNAs were always found in association with La France disease (Harmsen *et al.,* 1989; Morten and Hicks, 1992).

Amino acid sequences have now been deduced from the nucleotide sequences of five dsRNAs of ABV1: L1 and L5 (this paper) and L3, M1, and M2 (Harmsen *et al.,* 1991). Only the L1 dsRNA shows homology to proteins in current protein databases. It encodes an RdRp based on the presence of the consensus sequences found in other RdRp's. Best matching sequences occur in dsRNAs of viruses of the ascomycete *S. cerevisiae,* the heterobasidiomycete *Ustilago maydis,* and the protozoans *Leishmania braziliensis* subsp. *guyanensis* and *Giardia lamblia.* Contrary to ABV1, these viruses have all their genetic information present on one segment of dsRNA, encoding an RNA-dependent RNA polymerase and a capsid protein, and are classified as Totiviridae. Other

fungal viruses contain two and sometimes three dsRNA segments encoding at least a polymerase and a capsid protein and are called Partitiviridae (Buck, 1986). ABV1 cannot be considered a member of either the Totiviridae or the Partitiviridae as its genetic information is present on more than three separate dsRNA molecules.

As anticipated by Van der Lende *et al.* (1994), the minor protein of M_r 115K present in 34-nm virions proved to be the RdRp encoded by L1 dsRNA. The deduced amino acid sequence predicts a somewhat larger protein (M_r 121K). Amino acid sequencing of the abundant M_r 90K protein showed it to be encoded by the L3 dsRNA; although the deduced amino acid sequence predicts a protein of M_r 87K, L3 dsRNA could also be translated *in vitro* into a M_r 90K protein (not shown). The isoleucine (I) residue at position 781 in the deduced amino acid sequence, as determined by Harmsen *et al.* (1991), turned out to be a valine (V) in the analyzed protein. This can be the result of a single base mutation. The largest virion-associated protein (M_r 120K) is probably encoded by the L2 dsRNA, the only remaining candidate dsRNA with sufficient coding capacity. In fact, the L2 dsRNA could be translated *in vitro* into a protein of M_r 120K (not shown) but at this point there is no additional evidence to assign the 120K virion-associated protein to the L2 dsRNA segment.

Of the six dsRNAs that are always associated with La France disease symptoms, the smallest is the M2 segment. This segment appears to encode a protein of M_r 38K (Harmsen *et al.,* 1991) which was found in protein extracts of diseased *A. bisporus* fruit bodies but not in isolated virions. The function of this protein is unknown, but it may be related to disease development. However, the deduced amino acid sequence does not resemble any known protein. A first step in the elucidation of its role could be the cellular localization of this protein using the antiserum raised against the *E. coli*-expressed M2 protein.

In the past, 34-nm virus particles have also been found in apparently healthy fruit bodies (Van Zaayen, 1979; Frost and Passmore, 1980), but these might have been collected from infected mushroom trays. Apparently, fruit bodies can be infected in the absence of malformation. The main symptom of La France disease is loss of crop (Van Zaayen, 1979); in localized areas the mycelium does not permeate the casing layer (a layer of soil, covering the full-grown compost) or ceases initial growth into the casing layer. The malformations arising during fruit body ontogeny may thus be related to active replication of the virus rather than its mere presence. One could speculate that the differences in codon usage between ABV1 genes and the genes of the host *A. bisporus* (Table 1) indicate that the virus was recently acquired by *A. bisporus.* These differences in codon usage may also be of significance in disease development. During virus replication efficient translation of viral mRNAs could lower the concentration of rare amino acyl tRNAs of *A. bisporus* below that needed for translation of its own mRNAs. Particularly the synthesis of *A. bisporus* proteins needed in high quantities during fruiting and substrate colonization, such as hydrophobins (Wessels, 1994) which are abundant in *A. bisporus* (L. G. Lugones and J. G. H. Wessels, unpublished), could be hampered. This might cause at least some of the symptoms associated with La France disease such as decreased ability to colonize the substrate and the formation of fewer and malformed fruit bodies.

REFERENCES

Argos, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Res.* 16, 9909–9916.

Bruenn, J. A. (1993). A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. *Nucleic Acids Res.* 21, 5667–5669.

Buck, K. W. (1986). Fungal virology: An overview. *In* "Fungal Virology" (K. W. Buck, Ed.), pp. 1–85. CRC Press, Boca Raton, FL.

Deahl, K. L., San Antonio, J. P., and Civerolo, E. L. (1987). Electrophoretic analysis of double-stranded RNA in stocks of cultivated mushroom (*Agaricus brunescens*). *Plant Dis.* 71(5), 430–433.

Frost, R. R., and Passmore, E. L. (1980). Mushroom viruses: A re-appraisal. *Phytopathol. Z.* 98, 272–284.

Furfine, E. S., and Wang, C. C. (1990). Transfection of the *Giardia lamblia* double-stranded RNA virus into *Giardia lamblia* by electroporation of a single-stranded RNA copy of the viral genome. *Mol. Cell. Biol.* 10, 3659–3662.

Goodin, M. M., Schlagnhaufer, B., and Romaine, C. P. (1992). Encapsidation of the La France disease-specific double stranded RNAs in 36-nm isometric viruslike particles. *Phytopathology* 82, 285–290.

Harmsen, M. C., Van Griensven, L. J. L. D., and Wessels, J. G. H. (1989). Molecular analysis of *Agaricus bisporus* double-stranded RNA. *J. Gen. Virol.* 70, 1613–1616.

Harmsen, M. C., Tolner, B., Kram, A., Go, S. J., De Haan, A., and Wessels, J. G. H. (1991). Sequences of three dsRNAs associated with La France disease of the cultivated mushroom (*Agaricus bisporus*). *Curr. Genet.* 20, 137–144.

Harmsen, M. C., Schuren, F. H. J., Moukha, S. M., Van Zuilen, C. M., Punt, P. J., and Wessels, J. G. H. (1992). Sequence analysis of the glyceraldehyde-3-phosphate dehydrogenase genes from the basidiomycetes *Schizophyllum commune, Phanerochaete chrysosporium* and *Agaricus bisporus. Curr. Genet.* 22, 447–454.

Hicks, R. T. G., and Haughton, K. L. (1986). Detection of double-stranded RNA in shake-cultures of *Agaricus bisporus* affected by 'La France disease.' *Trans. Br. Mycol. Soc.* 86, 578–584.

Higgins, D. G., and Sharp, P. M. (1988). Clustal: A package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237–244.

Icho, T., and Wickner, R. B. (1989). The double stranded RNA genome of yeast virus L-A encodes its own putative RNA polymerase by fusing two open reading frames. *J. Biol. Chem.* 264, 6716–6723.

Koonin, E. V., and Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acids. *Crit. Rev. Biochem. Mol. Biol.* 28(5), 375–430.

Koons, K., Schlagnhaufer, B., and Romaine, C. P. (1989). Double stranded RNAs in mycelial cultures of *Agaricus bisporus* affected by La France disease. *Phytopathology* 79, 1272–1275.

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* 227, 1435–1441.

Marino, R., Saksena, K. N., Schuler, M., Mayfield, J. E., and Lemke, P. A. (1976). Double-stranded ribonucleic acid in *Agaricus bisporus. Appl. Environ. Microbiol.* 31, 433–438.

Morris, T. J., and Dodds, J. A. (1979). Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. *Phytopathology* 69, 854–858.

Morten, K. J., and Hicks, R. T. G. (1992). Changes in dsRNA profiles in *Agaricus bisporus* during subculture. *FEMS Microbiol. Lett.* 91, 159–163.

Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 8, 3867–1989.

Revill, P. A., Davidson, A. D., and Wright, P. J. (1994). The nucleotide sequence and genome organization of mushroom bacilliform virus: A single-stranded RNA virus of *Agaricus bisporus* (Lange) Imbach. *Virology* 202, 904–911.

Ribas, J. C., and Wickner, R. B. (1992). RNA-dependent RNA polymerase consensus sequence of the L-A double-stranded RNA virus: Definition of essential domains. *Proc. Natl. Acad. Sci. USA* 89, 2185–2189.

Ross, R. C., Brown, G. A., and Romaine, C. P. (1986). Recent experience in detecting viral double-stranded RNA in commercial mushroom crops and its effect on yield. *In* "Cultivating Edible Fungi" (P. J. Wuest, D. J. Royse, and R. B. Beelman, Eds.), pp. 321–329. Elsevier, Amsterdam.

Sanger, F., Nicklin, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

Tavantzis, S. M., Romaine, C. P., and Smith, S. H. (1980). Purification and partial characterization of a bacilliform virus of *Agaricus bisporus:* A single stranded RNA mycovirus. *Virology* 105, 94–102.

Tavantzis, S. M., Romaine, C. P., and Smith, S. H. (1983). Mechanism of genome expression in a single stranded RNA virus from the cultivated mushroom [*Agaricus bisporus* (Lange). Imbach]. *Phytopathol. Z.* 106, 45–50.

Van der Lende, T. R., Harmsen, M. C., and Wessels, J. G. H. (1994). Double-stranded RNAs and proteins associated with the 34 nm virus particles of the cultivated mushroom *Agaricus bisporus. J. Gen. Virol.* 75, 2533–2536.

Van der Vlugt, R. A. A., Schaap, P. J., Müller, Y., Thurston, C. F., Sonnenberg, A. S. M., Visser, J., and Van Griensven, L. J. L. D. (1993). A codon usage table for *Agaricus bisporus. Cultiv. Mushroom Res. Newsl.* 1, 50–52.

Van Zaayen, A. (1979). Mushroom viruses. *In* "Viruses and Plasmids in Fungi" (P. A. Lemke, Ed.), pp. 239–324. Dekker, New York.

Wach, M. P., Sriskantha, A., and Romaine, C. P. (1987). Double-stranded RNAs associated with La France disease of the commercial mushroom. *Phytopathology* 77, 1321–1325.

Wang, A. L., and Wang, C. C. (1986). Discovery of a specific double-stranded RNA virus in *Giardia lamblia. Mol. Biochem. Parasitol.* 21, 269–276.

Wessels, J. G. H. (1994). Developmental regulation of cell wall formation. *Annu. Rev. Phytopathol.* 32, 413–437.

White, T. C., and Wang, C. C. (1990). RNA dependent RNA polymerase activity associated with the double-stranded RNA virus of *Giardia lamblia. Nucleic Acids Res.* 18, 553–560.

Widmer, G., and Patterson, J. L. (1991). Genomic structure and RNA polymerase activity in *Leishmania* virus. *J. Virol.* 65, 4211–4215.

Widmer, G., Comeau, A. M., Furlong, D. B., Wirth, D. F., and Patterson, J. L. (1989). Characterization of an RNA virus from the parasite *Leishmania. Proc. Natl. Acad. Sci. USA* 86, 5979–5982.